

Location Prediction on Twitter Using Machine Learning Techniques

P. Manish¹, N. T. Priyanka², Ch. Ushaswini³, D. Sravanthi⁴, G. Venkat Sai⁵
^{1,2,3,4,5}Member, Pragati Engineering College

Abstract - Location prediction of users from online social media brings considerable research these days. Automatic recognition of location related with or referenced in records has been investigated for decades. As a standout amongst the online social network organization, Twitter has pulled in an extensive number of users who send a millions of tweets on regular schedule. Because of the worldwide inclusion of its users and continuous tweets, location prediction on Twitter has increased noteworthy consideration in these days. Tweets, the short and noisy and rich natured texts bring many challenges in research area for researchers. In proposed framework, a general picture of location prediction using tweets is studied. In particular, tweet location is predicted from tweet contents. By outlining tweet content and contexts, it is fundamentally featured that how the issues rely upon these text inputs. In this work, we predict the location of user from the tweet text exploiting machine learning techniques namely naïve bayes, Support Vector Machine and Decision Tree.

Index Terms - online social network organization, Tweets, naïve bayes, Support Vector Machine and Decision Tree.

I. INTRODUCTION

Users may post explicitly their location on the tweet text they post, whereas in certain cases the location may be available implicitly by including certain relevant criteria. Tweets are not a strongly typed language, in which users may post casual with emotion images. Abbreviated form of text, misspellings, and extra characters of emotional words makes tweet texts noisy. The techniques applied for normal documents are not suited for analysing tweets. The character limitations of tweets about 140 characters may make the tweet uneasy to understand if the tweet context is not studied.

The issue of location prediction related named as geo location prediction is examined for Wikipedia and web page documents. Entity recognition from these formal

documents has been researched for years. Different types of content and context handling on these documents are also studied extensively. However, the location prediction problem from twitter depends highly on tweet content. Users living in specific regions, locations may examine neighborhood tourist spots, landmarks and buildings and related events.

II. PROCEDURE FOR PAPER SUBMISSION

A. Review Stage

Submit your manuscript electronically for review. prepare it in two-column format, including figures and tables (until it doesn't fit properly and data is not visible).

B. Final Stage

After your paper has been accepted. The authors of the accepted manuscripts will be given a copyright form and the form should accompany your final submission.

C. Figures

As said, to insert images in Word, position the cursor at the insertion point and either use Insert | Picture | From File or copy the image to the Windows clipboard and then Edit | Paste Special | Picture (with —Float over text unchecked).

III. EXISTING SYSTEM

In the Existing system to the problem of finding location from social media content. The Social Networks from and motivated by Term frequency (TF) and inverse document frequency (IDF), they arrived Inverse City Frequency (ICF) and Inverse Location Frequency (ILF) respectively. They raked the features by using these frequency values and TF then by TF values. From this they arrived those local words spread in document in few places and have high ICF and ILF values. They approached model for identifying local words indicative or used in certain locations only.

They aimed to identify automatically by ranking the local words by their location, and they find their degree of association of location words associated to particular location or cities.

IV. PROPOSED SYSTEM

Live stream of twitter data is collected as dataset using authentication keys. The aim of proposed system is to predict the user location from twitter content considering user home location, tweet location and tweet content. To handle this we used three machine learning approaches to make prediction easier and finding the best model amongst them. Live tweet stream from twitter for keyword “apple” is collected and stored in Tweetable. Live twitter data can be collected by registering a consumer_key, consumer_secret, access_token, access_token_secret for authentication and collecting live stream of tweets. We have collected more than 1000 tweets of particular keywords such as Indian city hashtag names. You can also search tweets based on hashtags.

V. REQUIREMENT SPECIFICATION

Software Requirements:

1. Python,
2. Django

Operating Systems:- Windows 10 64 bit OS

Technologies and Languages used to Develop

1.Python

Debugger and Emulator: Any Browser (Particularly Chrome) Hardware Requirements

Hardware Requirements:

Processor: Intel i3

RAM: 4 GB

Space on Hard Disk: minimum 1 TB VI.

PUBLICATIONPRINCIPLES

VI. FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For

feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are,

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

ECONOMICAL FEASIBILITY

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

VII. SYSTEM DESIGN

SYSTEM ARCHITECTURE:

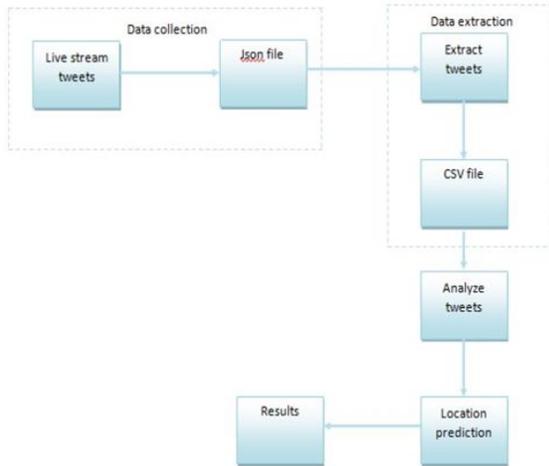


Figure 1

DATA FLOW DIAGRAM:

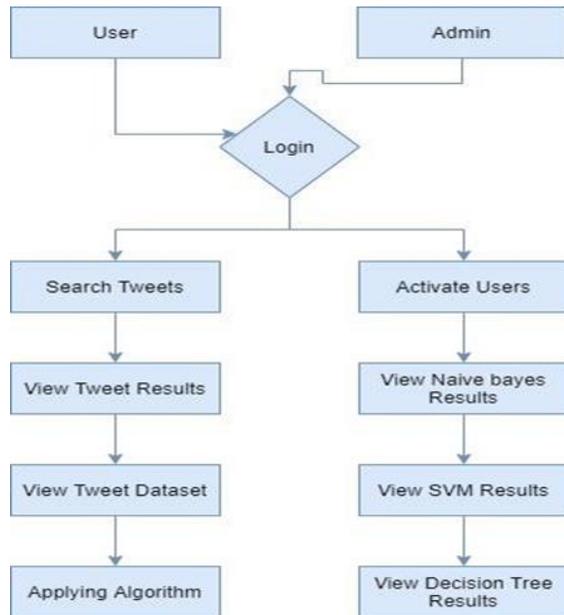


Figure 2

VII. CONCLUSION

Three locations are considered from twitter data, namely home location, mentioned location and tweet location. When the twitter data is considered, geolocation prediction becomes a challenging problem. The tweet text nature and number of characters limitation make it hard to understand and analyze. In this work, we have predicted the geolocation of user from their tweet text using machine learning algorithms. We have implemented three algorithms to show the better performed one,

which is suitable for geolocation prediction problem. Our experiment analysis concluded that decision tree is suitable for tweet text analysis and location prediction problem.

REFERENCES

- [1] Han, Bo & Cook, Paul & Baldwin, Timothy. (2012). Geolocation Prediction in Social Media Data by Finding Location Indicative Words. 24th International Conference on Computational Linguistics - Proceedings of COLING 2012: Technical Papers. 1045-1062.
- [2] Ren K., Zhang S., Lin H. (2012) Where Are You Settling Down: Geo-locating Twitter Users Based on Tweets and Social Networks. In: Hou Y., Nie JY., Sun L., Wang B., Zhang P. (eds) Information Retrieval Technology. AIRS 2012. Lecture Notes in Computer Science, vol 7675. Springer, Berlin, Heidelberg.
- [3] Han, Bo & Cook, Paul & Baldwin, Timothy. (2014). Text-Based Twitter User Geolocation Prediction. The Journal of Artificial Intelligence Research (JAIR). 49. 10.1613/jair.4200.
- [4] Li, Rui & Wang, Shengjie & Chen-Chuan Chang, Kevin. (2012). Multiple Location Profiling for Users and Relationships from Social Network and Content. Proceedings of the VLDB Endowment. 5. 10.14778/2350229.2350273.
- [5] Jalal Mahmud, Jeffrey Nichols, and Clemens Drews. 2014. Home Location Identification of Twitter Users. ACM Trans. Intell. Syst. Technol. 5, 3, Article 47 (July 2014), 21 pages. DOI: <http://dx.doi.org/10.1145/2528548>
- [6] Miura, Yasuhide, Motoki Taniguchi, Tomoki Taniguchi and Tomoko Ohkuma. "A Simple Scalable Neural Networks based Model for Geolocation Prediction in Twitter." NUT@COLING (2016).
- [7] A. Schulz, A. Hadjakos, H. Paulheim, J. Nachtwey, and M. M" uhlh" auser, "A multi-indicator approach for geolocalization of tweets," in Proc. 7th Int. Conf. on Weblogs and Social Media, 2013.
- [8] R. Li, S. Wang, H. Deng, R. Wang, and K. C.-C. Chang, "Towards social user profiling: unified and discriminative influence model for inferring home locations," in Proc. 18th ACM Int. Conf. on

Knowledge Discovery and Data Mining, 2012, pp. 1023–1031.

- [9] B. Han, P. Cook, and T. Baldwin, “A stacking-based approach to twitter user geolocation prediction,” in Proc. 51st Annual Meeting of the Association for Computational Linguistics System Demonstrations, 2013, pp.
- [10] D. Flatow, M. Naaman, K. E. Xie, Y. Volkovich, and Y. Kanza, “On the accuracy of hyper-local geotagging of social media content,” in Proc. 8th ACM Int. Conf. on Web Search and Data Mining, 2015, pp. 127–136.
- [11] O. V. Laere, J. A. Quinn, S. Schockaert, and B. Dhoedt, “Spatially aware term selection for geotagging,” IEEE Trans. Knowl. Data Eng., vol. 26, no. 1, pp. 221–234, 2014.
- [12] J. Mahmud, J. Nichols, and C. Drews, “Where is this tweet from? inferring home locations of twitter users,” in Proc. 6th Int. Conf. on Weblogs and Social Media, 2012.